THESIS
AN EVALUATION OF TEACHERS' ASSIGNED GRADES
IN MUSIC AND OF THE INFLUENCE OF OUT-OF-SCHOO
MUSIC STUDY IN TERMS OF RESULTS OF THE
KWALWASSER-RUCH TEST OF MUSICAL ACCOMPLISHMEN
Elizabeth Virginia Carter
Master of Music Education
1947

AN ABSTRACT OF THE

THESIS

An Evaluation of Teachers' Assigned Grades in Music and of the Influence of Out-of-School Music Study in Terms of Results of the Kwalwasser-Ruch Test of Musical Accomplishment

by

Elizabeth Virginia Carter

This thesis has two objectives. They are, namely:

(a) To test the validity of teachers' assigned grades in music as compared with the results of a standardized test of musical accomplishment.

(b) To determine the influence, if any, of out-of-school music study on pupils' musical accomplishment as shown by a standardized test.

In the analysis of the problem our plan involved several steps. They are:

1. Examination of the literature for information regarding available achievement type music tests.
2. Selection of the test. The Kwalwasser-Ruch Test of Musical Accomplishment, a 1923 publication, has been and is one of several worthwhile contributions to the testing phase of music education.
3. More detailed study of the literature pertaining to the Kwalwasser-Ruch Test. The literature is extensive due to the age of the test.
4. Obtain subjects for testing. Subjects were obtained through the cooperation of the supervisors of Music in the Melrose, Brockton, and Needham schools.
5. Administering and scoring the tests.

   In the Brockton and Melrose schools, the author of the research problem administered the test. In the Needham school, the music supervisor administered the test. The tests were hand scored.
6. The assembly of teachers' assigned grades for the pupils being tested involved a reconciliation of three different grading systems and necessitated a method whereby these categorical ratings

were transformed into numerical scores suitable for handling. This material is discussed fully in the body of the thesis, page 21.

7. The length of time spent in out-of-school music study was assembled from material furnished on the test booklet. This music study was computed in half-hour lessons for the total period of study.

Solution of the problem was approached with a sampling which included 207 junior high school girls and boys from Needham, Melrose, and Brockton public schools. Grades, seven-65 pupils, eight-31 pupils, nine-111 pupils were represented.

The scores of the 207 pupils on the Kwalwasser-Ruch Test will be correlated with the teachers' grades for the first semester and with the number of half-hours for the total period of instruction in music outside of school.

Data used in the presentation of the problem took the form of tables and correlations which pertained to each of the relationships referred to in the two objectives of the thesis.

We assumed, in the study, that every pupil in each grade had an equal or equivalent number of years of school music study. We also assumed that the test does not measure all phases of achievement in music, but does measure a representative number of basic knowledges, namely:

Recognition of Syllable Names, Knowledge of Musical Symbols and Terms, Detection of Pitch Errors in a Familiar Melody, Detection of Time Errors in a Familiar Melody, Recognition of Pitch Names, Knowledge of Time Signatures, Knowledge of Key Signatures, Knowledge of Note Values, Knowledge of Rest Values, and Recognition of Familiar Melodies from Notation. (These knowledges are titles of the sub-tests in the Kwalwasser-

Ruch Test of Musical Accomplishment.

Instead of listing the hypothetical conclusions with their implications, as has been done in the thesis, we shall give the reader our findings and conclusions which allude to the hypothetical conclusions.

We assumed that the validation of the Kwalwasser-Ruch Test was one of face validation, there being no item analysis data offered in the manual to substantiate any other supposition.

Reliability data which we find in Table 1 of the Appendix, page 31, are sufficient in evaluating the practical significance of test results. Reliabilities as revealed by the standard error were such that use of the test seemed feasible.

In the total sample, the mean for the group is 142.39 as compared with 156.0 for the national group. The standard deviation for the sample was 43.3 as compared with 51.5 for the national group. The mean figure indicates to us that there are 123 persons out of 207 who are in and above the interval in which we find the score 142.39. This is approximately 54% of the tested population or a little more than half of the sample. The standard deviation is in reality a standard with which deviations of individual scores may be compared. According to our figures, scores in the total sample will vary by ± 43.3 . A standard deviation of 43.3 would seem unusually high were it nor for the fact that the range of scores is also great, that is 194.

In the seventh grade group, the mean more nearly approaches the mean of the national norms. In fact, its mean of 152.46 represents a score which, according to the percentile chart prepared by the test authors, is at the eightieth percentile. In the eighth grade group, the mean is

Digitized by the Internet Archive
in 2015

https://archive.org/details/evaluationofteac00cart

165.65 and represents a score which is also above the mean for the national group. The ninth grade group does not show as creditably in relation to the national norms, in that the mean of the sample population in that grade was 129.99 as compared with 156 for the norm group.

The same explanation obtains in regard to the eighth and ninth grades' mean scores as applies to the explanation for the mean of the total group and the seventh grade. Specific figures vary from group to group, however.

The coefficient of correlation between teachers' grades and pupils' scores on the Kwalwasser-Ruch Test is positive but not high enough to be regarded as significant. (The coefficient of correlation was .24) We may not predict future success or achievement on the basis of our findings, for, though statistically significant, the coefficient of correlation must approach approximately .65 for use in prediction.

The trend of evidence would lead one to believe that the original hypothetical conclusions in the outline of research may have some credence. The evidence points either to a marking system which is inadequate or to one which places emphasis on varying phases of school music work. It might also be implied that some phases of subject matter involved in teacher judgments were not receiving the emphasis which would demonstrate itself in a test designed to measure basic knowledges. If we are to believe the results which present themselves, there is either some poor teaching or there are some areas of knowledge which are left untouched. Any one test, regardless of the areas it measures, could not decide that the quality of teaching or quality of students' work was poor. It could, however, point out that weaknesses exist, and it would serve as a point of departure for diagnosis of pupil performance in a

more careful manner.

In the results of the correlation between outside study and test scores, the coefficient of correlation was .88, indicating that there is a high positive relationship between the two. By inspection, while correcting the test papers, similar results were noticable, although more than 100 of the 207 testees had seven hours or less of out-of-school music study. Let us then conclude, that in the group studied, outside instruction has supplemented classroom instruction to such an extent that at least half of the pupils in the group who have had out-of-school music study have made scores better than the mean score for their grade group.

If our original hypothesis postulates higher scores for the pupils who have had private instruction and our data show that some pupils who have received such instruction have earned scores below the mean, we may conclude one of two things. We may conclude that either our hypothesis is inaccurate or that the type and quality of instruction which the pupils receive in their out-of-school study must vary greatly. We should ask ourselves, " Why the latter explanation?"

Subject matter , as taught by various individuals, will have varying emphases. Assuming the average intelligence of pupils, this variation of emphasis will project itself into and influence pupil performance--whether in a test or in some other form of response to instructional activity. The negative response to private instruction might indicate a lack of aptitude to absorb information satisfactorily, no matter what the specific stimulus or situation. This particular phase of our findings furnishes possibilities for additional research.

BOSTON UNIVERSITY

COLLEGE OF MUSIC

Thesis

An Evaluation of Teachers' Assigned Grades in Music and of the Influence of Out-of-School Music Study in Terms of Results of the Kwalwasser-Ruch Test of Musical Accomplishment

by

Elizabeth Virginia Carter

B.S., Mus.Ed., Southern University, 1942

submitted in partial fulfillment of the
requirements for the degree of
Master of Music Education

1947

Approved by

First Reader . . . . Kenneth G. Kelley . . . .

Dean of the College of Music

Second Reader . . . . Keith D. Snyder . . . .

Assistant Professor of Music Education

TABLE OF CONTENTS

Page

I. Statement of Purpose of Investigation 3

II. Plan of Research 4

III. Body of the Thesis 9

Introduction 10

Frequency Table for Scores of Total Group 11

Frequency Table for Scores of Grade Seven 12

Frequency Table for Scores of Grade Eight 13

Frequency Table for Scores of Grade Nine 14

Table of Sub-Test Scores 15

Table of Percentile Rank of Sample in Relation to Rank of Norm Group 17

Tabulation of Out-of-School Music Study 19

Table Indicating T Scores obtained from Transforming Letter Grades 20

Transforming Letter Grades into Numerical Scores 21

Correlations
Use of Pearson Product-Moment by Long Method and Durost-Walker Correlation Chart 23

IV. Findings and Conclusions 24

V. Appendix

Pertinent Information About the Kwalwasser-Ruch Test of Musical Accomplishment 28

Table 1- Reliability Coefficients 31

Table 2- Percentile Scores for Norm Group 32

VI. Bibliography 33

STATEMENT OF PURPOSE OF INVESTIGATION

This thesis has two objectives. They are, namely:

(a) To test the validity of teachers' assigned grades in music as compared with the results of a standardized test of musical accomplishment.

The validity of this research problem may be questioned because of the subjectiveness which is introduced by the use of teachers' grades as one of the variables in the study. Teachers' grades, though subjective in nature, are accepted forms of evaluation of pupil achievment. This acceptability makes them feasible for use in a validation study. The use of a standardized commercial test of musical accomplishment gives us a "control mechanism" against which we can check the validity or lack of validity of teachers' assigned grades.

(b) To determine the influence, if any, of out-of-school music study on pupils' musical accomplishment as shown by a standardized test.

We believe that some learning, whether or not specifically related to the teaching programs in school, is carried over from the instruction which pupils receive in their study of music outside of school.

PLAN OF RESEARCH

The steps involved in analysis of the problem were these:

1. Examination of the literature for information regarding available achievement type music tests.
2. Selection of the test. The Kwalwasser-Ruch Test of Musical Accomplishment, which has been in existence since 1923, remains one of several worthwhile contributions to the testing phase of music education. There was a need for a test which had been in use sufficiently long so that it was more familiar to supervisors of music than some of the more recent and perhaps less frequently used tests. The word accomplishment in the title may be misleading, for all phases of music accomplishment are not tested. However, the test questions embrace many of the basic knowledges taught in school music programs. Thus the Kwalwasser-Ruch Test of Musical Accomplishment was my choice.
3. More detailed study of the literature pertaining to the Kwalwasser-Ruch Test. The literature is extensive because the test is about twenty-four years old.
4. Obtain subjects for testing.
   Subjects were obtained through the cooperation of the supervisors of music in the Melrose,Brockton, and Needham schools.
5. Administering and scoring the tests.
   In the Brockton and Melrose schools, the author of the research problem administered the test. In the Needham school the music supervisor administered the test. The tests were hand scored.

6. The assembly of teachers' assigned grades for the pupils being tested involved a reconciliation of three different grading systems and necessitated a method whereby these categorical ratings or letter grades were transformed into numerical scores suitable for handling. The topic "Transforming Letter Grades into Numerical Scores" on page 21 of the body of the thesis will explain more in detail the necessity of such a procedure and the steps involved.
7. The length of time spent in out-of-school music study was assembled from material furnished on the test booklet. This study was computed in half-hour lessons for the total period of study. The data was assembled in units of two weekly half-hour lessons or their equivalent.

Solution of the problem was approached with a sampling which included 207 junior high school girls and boys from three suburban public schools. Of this number there were 65 in the seventh grade, 31 in the eighth grade, and 111 in the ninth grade.

In the Needham Junior High School, an experimental group from grades seven, eight, and nine was used. The group was selected to include pupils of varying abilities because music classes in this school are organized on the basis of homogeneous grouping.

In the Melrose school one seventh grade was available.

In the Brockton schools, two ninth grade classes- one a group in which there were pupils who were average and below average in their level of general school achievement and another group with a more heterogeneous population.

The scores of the 207 pupils on the Kwalwasser-Ruch Test will be

correlated with the teachers' grades for the first semester and with the number of half-hours for the total period of instruction in music outside of school.

For the analysis of the two relationships we have set up in our problem we used several kinds of data which were similar for each relationship. The reliability and validity data of the Kwalwasser-Ruch Test concern both relationships. This material was furnished by the test manual and is reproduced in the Appendix, page 28 . Pupils' test scores pertain to both relationships and were obtained from the test booklets. The pupils' assigned grades for use in the correlation between teachers' grades and the test scores were obtained from the teachers' records. The length of time spent in music study outside of school was computed in half-hours for the total period of study. This information or its equivalent (as was mentioned in step 7 of the analysis) was furnished by the pupils on the outside of the test booklets.

The data was tabulated to indicate the total number of pupils who were tested and the letter grades which were assigned to them for use in the relationship between pupils' test scores and teachers' assigned grades.

Data was also tabulated to indicate the number of pupils who have studied music outside of school and the number of half-hour lessons per total period of study for use in connection with the relationship between pupils' test scores and out-of-school music study.

Further tabulation included the total and sub-test scores from which certain statistical data were obtained (Mean and Standard Deviation). The final statistical data obtained were the coefficients of correlation

between teachers' grades and pupils' test scores and between pupils' test scores and the amount of out-of-school music study.

We assumed, in this study, that every pupil in each grade had an equal or equivalent number of years of school music study. We also assumed that the test does not measure all phases of achievement in music, but does measure a representative number of basic knowledges, namely:

1. Knowledge of Musical Symbols and Terms
2. Recognition of Syllable Names
3. Detection of Pitch Errors in a Familiar Melody
4. Detection of Time Errors in a Familiar Melody
5. Recognition of Pitch Names
6. Knowledge of Time Signatures
7. Knowledge of Key Signatures
8. Knowledge of Note Values
9. Knowledge of Rest Values
10. Recognition of Familiar Melodies from Notation

1. We conclude, hypothetically, that there is a positive correlation between teachers' grades and pupils' Kwalwasser-Ruch test scores.
2. We also conclude, hypothetically, that there is a positive correlation between pupils' test scores and the factor of out-of-school music study.

The results of a study such as this one may have several implications. Among these are, namely:

1. If results indicate that correlation between teachers' grades and pupils' test scores is negligible, it may mean that:

   a. There may be a need for refinement of teachers' grades, or that
   b. Areas of subject matter emphasis may need revision, or that,
   c. Studies should be conducted using other music achievement tests to see if correlations obtained in this study are relatively the same.

2. If results indicate that correlation between teachers' grades and pupils' scores is high, it may mean that:
   a. The teachers' grading methods represent a fairly accurate summation of pupil achievement in the basic knowledges.
   b. The test material and the teaching material used in instructing the tested group are similar.
3. If results indicate that the correlation between pupils' scores and out-of-school music study is negligible, we may conclude that the material taught in the classroom is not related to the material used in out-of-school music study and vice versa. We may also conclude that the manner of presentation and emphasis in instruction in the school music classroom varies such that no relationship can be established between the two.
4. If results indicate that the correlation between pupils' test scores and out-of-school music study are high we may conclude that out-of-school music instruction has supplemented classroom instruction.

BODY OF THE THESIS

This thesis has as its main body of material certain frequency tables which pertain to scores of the total test group, scores of grades seven, eight, and nine, and more simply organized tables which have to do with pupils' sub-test scores, the percentile rank of the sample in relation to the rank of the norm group, the number of hours of out-of-school music study of the pupils, transformation of letter grades into numerical scores, and a table indicating T-scores obtained from the transformation of letter-grades. In addition to the above-named tables, there is information pertinent to the actual computation of the coefficients of correlation between the pupils' test scores and teachers' grades, and between the pupils' test scores and the amount of out-of-school music study.

The frequency tables and the other statistical tables which use intervals were set up with step intervals. This is true of the first four frequency tables and the table of half-hours of out-of-school music study. The first four frequency tables were set up for the purpose of computing the mean and standard deviation.

A key to the symbols used in this section may be helpful to the reader and is included here.

1. r represents the coefficient of correlation computed for an observed sample.
2. Σ is the sign of summation; a symbol which indicates that certain quantities represented by a symbol or symbols following the Σ are to be added. Σxy may be read "The sum of

x times y."

3. x represents a deviation of a score from the mean of its distribution.
4. y is used in similar fashion as x when there is more than one factor.
5. N equals the number of cases under consideration.
6. f represents a frequency.
7. X represents a gross score.
8. cf represents a cumulative frequency.
9. σ represents a standard deviation of a sample
10. i represents the width of a step-interval.

A detailed explanation of transforming categorical ratings or letter grades to numerical scores and in turn converting these letter grades to "T" scores is given following the frequency tables and the other tables which have been mentioned in the first paragraph, page 9.

Frequency Table for the Scores of the Total Test Group

	f	x'	fx'	$f(x')^2$	cf
237-223	4	6	24	144	207
222-208	10	5	50	250	203
207-193	17	4	68	272	193
192-178	18	3	54	162	176
177-163	20	2	40	80	158
162-148	30	1	30	30	138
147-133	24	0			108
132-118	16	-1	-16	16	84
117-103	22	-2	-44	88	68
102-88	23	-3	-69	207	46
87-73	13	-4	-52	208	23
72-58	8	-5	-40	200	10
57-43	2	-6	-12	72	2

N equals 207

Using this formula, we find the mean:

Mean equals $M' + \frac{\Sigma fx'}{N} .1$

Mean equals 142.39

Using this formula, we find the standard deviation

σ equals $i\sqrt{\frac{\Sigma f(x)^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2}$

σ equals 43.28505 or 43.3

Frequency Table for Scores of Grade Seven

	f	x'	fx'	$f(x')^2$	cf
237-223	2	5	10	50	65
222-208	5	4	20	80	63
207-193	5	3	15	45	58
192-178	6	2	12	24	53
177-163	6	1	6	6	47
162-148	13	0			41
147-133	8	-1	-8	8	28
132-118	6	-2	-12	24	20
117-103	5	-3	-15	45	14
102-88	7	-4	-28	112	9
87-73	1	-5	- 5	25	2
72-58	1	-6	-6	36	1

N equals 65

Using the same formulae as on the previous page, we find the:

Mean, which equals 152.465 or 152.47

Standard deviation, which equals 39.6045 or 39.60

Frequency Table for Scores of Grade Eight

	f	x'	Fx'	$f(x)'^2$	cf Cf
237-223	1	4	4	16	31
222-208	1	3	3	9	30
207-193	6	2	12	24	29
192-178	4	1	4	4	23
177-163	5	0			19
162-148	6	-1	-6	6	14
147-133	1	-2	-2	4	8
132-118	4	-3	-12	36	7
117-103	3	-4	-12	48	3

N equals 31

Using the formulae given on the first frequency table, we find:

Mean, which equals 165.65

Standard Deviation, which equals 32.37

Frequency Table for the Scores in Grade Nine

	f	x'	fx'	$f(x')^2$	cf
237-223	1	7	7	49	111
222-208	4	6	24	144	110
207-193	6	5	30	150	106
192-178	8	4	32	128	100
177-163	9	3	27	71	100
162-148	11	2	22	44	83
147-133	15	1	15	15	72
132-118	6	0			57
117-103	14	-1	-14	14	51
102- 88	16	-2	-32	64	37
87- 73	12	-3	-36	108	21
72- 58	7	-4	-28	112	9
57- 43	2	-5	-10	50	2
	N equals 111		37	949	

Using the formulae which have been previously referred to, we find the

Mean, which is 129.99

Standard Deviation, which is 43.50

Table of Sub-Test Scores

Test 1

Score	N
25	6
24	4
23	10
22	12
21	11
20	22
19	32
18	23
17	20
16	15
15	16
14	12
13	8
12	3
11	4
10	5
9	2
8	1
1	1
	207

Test 2

Score	N
25	12
24	19
23	26
22	15
21	22
20	19
19	19
18	17
16	14
15	7
14	8
13	2
12	2
10	2
7	1
5	1
3	1
0	1
	207

Test 3

Score	N
25	57
20	55
15	34
10	28
5	22
0	11
	207

Test 4

Score	N
15	84
12	71
9	31
6	16
3	4
0	1
	207

Score	N
20	63
19	18
18	8
17	10
16	5
15	4
14	3
13	3
12	5
11	5
10	23
9	6
8	4
7	4
6	1
5	9
4	6
3	3
2	6
1	6
0	16
	207

The first number in each column indicates the highest possible score to be earned in each sub-test. This tabulation indicates how the scores were distributed among the 207 testees. Scores vary because some sub-tests allowed 5 points for each correct answer, while others allowed 3, and still others allowed one.

Test 6		Test 7		Test 8		Test 9		Test 10	
Scores	N	Scores	N	Scores	N	Scores	N	Scores	N
20	17	30	16	15	55	15	25	50	17
18	22	28	6	12	61	12	41	45	10
16	21	24	7	9	64	9	51	40	20
14	30	22	6	6	15	6	36	35	5
12	35	20	11	3	34	3	34	30	15
10	35	18	11	0	8	0	20	25	25
8	20	16	8					20	24
6	11	14	9					15	21
4	9	12	5					10	24
2	7	10	7					5	26
		8	7					0	20
		6	9						
		4	18						
		2	39						
		0	41						

A Table Indicating the Percentile Rank of the Sample Population in Relation to the Norm Group's Percentile Rank

(Refer to Table 2 on Page 32)

Grade Seven	Needham	Melrose
90th percentile (and above)	15	9
80th percentile (")	5	6
70th percentile "	3	9
60th percentile "	1	4
50th percentile (Median)	1	2
40th percentile	2	4
30th percentile	0	3
20th percentile	1	0
10th percentile	1	1
Grade Eight		
90th percentile (and above)	8	
80th percentile "	5	
70th percentile "	8	
60th percentile "	2	
50th percentile (Median)	1	
40th percentile	4	
30th percentile	3	
20th percentile	0	
10th percentile	0	
Below 10th percentile	1	

Table of Percentile Rank (Con't.)

Grade Nine

	Brockton	Needham
90th percentile (and above)	0	1
80th percentile "	4	2
70th percentile "	8	2
60th percentile "	2	1
50th percentile (Median)	9	2
40th percentile	17	4
30th percentile	7	0
20th percentile	18	1
10th percentile	21	2
Below 10th percentile	0	0

The percentile ranks indicated here were set up and based on the scores of the norm group. We were able to locate the positions of pupils from each school sampling in relation to the norms for the national school population.

Tabulation of Out-of-School Music Study

Half-hours per total period of time studied	Number of pupils in sample who have studied
300-314	2
285-299	1
270-284	0
255-269	2
240-254	1
225-239	2
210-224	1
195-209	3
180-194	3
165-179	4
150-164	12
135-149	0
120-134	1
105-119	1
90-104	13
75- 89	7
60- 74	11
45- 59	14
30 -44	8
15 -29	11
0 -14	110
	207

The table indicates that more than half of the tested population has had less than 15 half-hours of music study outside of school.

Table Indicating T Scores Obtained from Transforming Letter Grades

	Letter Grade	Number	T Score
Brockton			
	A	31	59
	B	38	49
	C	20	40
	D	7	32
Melrose			
	A	18	44
	B	18	38
	C	2	27
Needham			
	A	19	60
	B	24	45
	C	28	34
	D	1	28
	F	1	24

These scores were used in computation of the coefficient of correlation between teachers' assigned grades and pupils' test scores.

TRANSFORMING LETTER GRADES INTO NUMERICAL SCORES

In a sample involving several teachers' letter grades some method must be used to transform these categorical ratings into numerical scores which can be more easily handled. In explanation and justification of the procedure involved we quote from Walker.[1]

> "It has been seen that few traits with which the research worker is concerned have units whose equality can be determined directly, Traits such as mental age, measures of capacity, the underlying ability "measured" by a paper and pencil test, etc., have units equal by definition. In still other situations individuals can be placed in an ordered series but establishing the equality of units even by definition appears dubious.
>
> If it seems reasonable to postulate a normal distribution of a trait and to assume that the ordered categories (A,B,C,D,F,) represent unequal intervals on the scale of that trait, we may utilize the relationship between the area and deviation for the normal curve to find numerical equivalents for the categories.
>
> Let us make the assumption that the ratings have a normal distribution and that these records have been made in large and unequal intervals. The persons rated "D", and below are all equally poor, but the only information that we have is that they constitute the lowest percent (an unknown quantity except in a specific problem) of the total group. This percent depends on number in the category and in the total group.
>
> By consulting a table of the normal curve we find that the lowest percent (unknown except in a specific problem) lies at a point which will be indicated in terms of a number of standard deviations below the mean. We go on from there to find the other division points in a similar manner.
>
> Obviously, however, it is not these division points which are needed to represent the typical positions of the various sub-groups, but some average measure. The problem is to find a satisfactory average value which can represent that sub group. That measure is the median of the sub-class which is easy to interpret. After finding the median values, we multiply them by 10 and add 50 producing what W.A. McCall has named T-scores."

[1] Walker, Helen M., Elementary Statistical Methods, 1943. Henry Holt and Company, pp. 182-183.

We will indicate the steps used in converting letter grades in hope that they may furnish additional information to the reader. Actual computation has been omitted and a table has been included which indicates letter grades assigned in each school, the number of pupils earning them and the resulting T score which was used in computation of the coefficient of correlation. The steps are as follows:

1. Set up a frequency distribution in the usual fashion.
2. Make a cumulative frequency distribution, adding all the cases below a certain given score and half of those at the score. Cumulation is to the midpoint of the interval.
3. Divide each cumulative frequency by N to produce a cumulative percentage.
4. Subtract 50% from each cumulative percentage. The result is the percentage of cases between ordinates at the given score and at the median.
5. Turn to a table of areas and abscissas of the normal curve and read from it the abscissa value of a point such that the given percentage of area lies between an ordinate at the point and an ordinate at the mean.
6. Multiply each abscissa value by 10 and add 50 to obtain the T-score."

CORRELATIONS

In obtaining the coefficient of correlation between the teachers' grades and pupils' test scores the Durost-Walker method of computing the Pearson Product moment was used. The use of the Durost-Walker Correlation Chart speeds up the calculation of the formula and reduces the working time considerably by several hours. There are several places for checks in the course of the procedure which insure accuracy.

In obtaining the coefficient of correlation between the pupils' test scores and the amount of out-of-school music study the long method of computing the Pearson Product moment was used. The formula is as follows:

$$r = \frac{\Sigma xy}{N\sqrt{\Sigma x^2 \Sigma y^2}}$$

r equals .88

The solution of the formula, as we see by the figures, results in a coefficient of correlation which equals .88.

FINDINGS AND CONCLUSIONS

We assumed that the validation of the Kwalwasser-Ruch Test was one of face validation, there being no item analysis data offered in the manual to substantiate any other supposition.

Reliability data which we find in Table 1 of the Appendix, page 31, are sufficient in evaluating the practical significance of test results. Reliabilities as revealed by the standard error were such that use of the test seemed advisable.

In the total sample, the mean for the group is 142.39 as compared with 156.0 for the national group. The standard deviation for the sample was 43.3, as compared with 51.5 for the national group. The mean figure indicates to us that there are 123 persons out of the 207 who are in and above the interval in which we find the score 142.39. This is approximately 54% of the tested population or a little more than half of the sample. The standard deviation is in reality a standard with which deviations of individual scores may be compared. According to our figures scores in the total sample will vary by ± 43.3. A standard deviation of this size would seem unusually large were it not for the fact that the range of scores is also great, i.e., 194.

In the seventh grade group, the mean more nearly approaches that of the national norms. In fact, its mean of 152.46 represents a score which, according to the percentile chart prepared by the test authors, is at the eightieth percentile. In the eighth grade group, the mean is 165.65 and represents a scores which is also above the mean for the national group. The ninth grade group does not show as creditably in its score in

relation to the national norms, in that the mean of the sample population in that grade was 129.99 as compared with 156 for the norm group.

The same explanation obtains in regard to the eighth and ninth grades' mean scores as applies to the explanation for the mean of the total group and the seventh grade. Specific figures vary from group to group, however. For example, in the eighth grade group we find that the mean figure of 165.65 represents a sizable increase over the mean score of the total group and over the median score of the eighth grade group in the national norms chart.(Table 2, Appendix, page 32.)

The coefficient of correlation between teachers' grades and pupils' scores on the Kwalwasser-Ruch Test is positive but not high enough to be regarded with significance.(The coefficient of correlation was .24). We may not predict future success or achievement on the basis of our findings for, though statistically significant, the coefficient of correlation must approach approximately .65 for use in prediction.

The trend of evidence would lead one to believe that the original hypothetical conclusions in the outline of research may have some credence. The evidence points either to a marking system which is inadequate or to one which places emphasis on varying phases of school music work. It might also be implied that some phases of subject matter involved in teacher judgements were not receiving the emphasis which would demonstrate itself in a test designed to measure basic knowledges. If we are to believe the results which present themselves, there is either some poor teaching or there are some areas of knowledge which are left untouched. Any one test, regardless of the areas it measures, couldnot decide that the quality of teaching or quality of students' work was bad.

It could, however, point out that weaknesses exist, and it would serve as a point of departure for diagnosis of pupil performance in a more careful manner.

In the results of the correlation between outside study and test scores, the coefficient of correlation was .88, indicating that there is a high positive relationship between the two. By inspection, while correcting the test papers, similar results were notable, although more than 100 of the 207 testees had 7 hours or less of out-of-school music study. Let us then conclude, that in the group studied,outside instruction has supplemented classroom instruction to such an extent that at least half of the pupils in the group who have had out-of-school music study have made scores better than the mean score for their grade group.

If our original hypothesis postulates higher scores on the basis of private instruction and our data show that some pupils who have received such instruction have earned scores below the mean, we may conclude one of two things. We may conclude that either our hypothesis is inaccurate or that the type and quality of instruction which the pupils receive in their out-of-school study must vary greatly. We should ask ourselves, "Why the latter explanation?"

Subject matter, as taught by various individuals, will have varying emphases. Assuming the average intelligence of pupils, this variation of emphasis will project itself into and influence pupil performance-whether in a test or in some other form of response to instructional activity. The negative response to private instruction might indicate a lack of aptitude to absorb information satisfactorily, no matter what

the specific stimulus or situation. This particular phase of our findings furnishes possibilities for additional research.

Dykema and Gehrkens[2] in Chapter XVI of their book "High School Music" furnish the views of two outstanding music educators on private instruction or what we have called out-of-school music study.

They suggest, among other things, a provision for credits, a more thorough type of instruction, and the genuine concern of music teachers for the constructive use of private instruction as a complement as well as a supplement to school music instruction.

> " We must see to it that our high school pupils are taught music and musicianship, and not merely manual dexterity.When properly taught, applied music is the most important single phase of music study because it exemplifies our basic principle, namely, that we learn by doing rather than by standing by and talking about, or even by listening.
>
> " The more enlightened private teaching we have, the better accompanists, finer instrumentalists, and more intelligent theory students shall we find in our high school groups.

In all, private instruction or out-of-school music study can be rewarding for all persons concerned, pupil, private teacher, the school teacher, the administrator and finally, but not least, the parent who pays the bills! A more thorough and a more well-regulated type of private instruction can pay dividends-- in a higher type of pupil performance.

[2] Dykema,Peter W. and Gehrkens,Karl W., The Teaching and Administration of High School Music, C.C. Birchard & Company, Boston, 1941. pp. 229,230,232,233,235.

APPENDIX

PERTINENT INFORMATION ABOUT
THE KWALWASSER-RUCH TEST OF MUSICAL ACCOMPLISHMENT

"The Kwalwasser-Ruch Test of Musical Accomplishment is designed to measure the achievement of pupils in the typical public school music course in the elementary and high school grades. The following list of titles of the separate tests of the Kwalwasser-Ruch Test of Musical Accomplishment indicate largely the separate factors which the test measures.

1. Knowledge of Musical Symbols and Terms
2. Recognition of Syllable Names
3. Detection of Pitch Errors in a Familiar Melody
4. Detection of Time Errors in a Familiar Melody
5. Knowledge of Time Signatures
6. Knowledge of Key Signatures
7. Recognition of Pitch Names
8. Knowledge of Note Values
9. Knowledge of Rest Values
10. Recognition of Familiar Melodies from Notation

The validity of the Kwalwasser-Ruch Test rests primarily upon the specifications adopted by the Music Supervisors' National Conference.[1] The report approved by the Educational Council outlined a course which contained the aims, materials, procedures, and attainments for each of the first eight grades. The test has been further checked against a number of courses of study in city school systems which have received national recognition for their work in public school music.

The reliability of the total scores and also of the scores on the ten separate tests as determined from 167, sixth, eighth, tenth, and twelfth grade pupils is shown in Table 1, page 31. The correlations were figured by the method of splitting the test items into chance halves,

---

[1] Bulletin No.1, 1921, Music Supervisors National Conference, at the Fourteenth Annual Meeting in St. Joseph, Mo., April 4 through 8, 1921.

i.e., correlating the sum of the points earned on the even-numbered items with the points earned on the odd-numbered items. The correlation (reliability coefficient) for the tests as a whole was found to be .97[2] .

A more meaningful way of stating this result is to say that the probable error of a score is approximately 6 score points, i.e., the chances are even that his true score falls between 68 and 92 and about twenty to one that his true score falls between 62 and 98.

Obviously, a pupil's ability in music is measured by the test in a much more reliable manner than by the ordinary final examination whose reliability seldom exceeds .50 to .75 in comparison with the .97 of the Kwalwasser-Ruch Test of Musical Accomplishment.

The norms reported in Table 2, page 32 are based upon scores earned by 5,414 pupils in grades IV to XII. The following cities are represented: Evanston, Illinois; Oakland, California; Rochester, New York; Mason City,Iowa; Pittsburgh, Pennsylvania; Denver, Colorado; and South Bend, Indiana.

The interpretation of the norms is quite simple. For example, if a pupil earns a score of 107 on the test, it will be seen by referring to the 50th percentile scores (Medians) in Table 2 on page 32 , that his accomplishment is above average for grades below the eighth, but below the accomplishment average for grades above and including the eighth. Whether the score is a good or poor showing depends upon the grade in which the pupils is located. In the fourth grade a score of 107 is superior, but in the high school it must be considered unsatisfactory. In table 2, the high school norms have been pooled since there is little

---

[2] 1.00 is considered perfect correlation.

growth in musical accomplishment in these grades. We must remember, however, that the pupils' scores in a selected population may vary considerably from the norm group depending upon whether or not the tested population has been "exposed" to similar teaching procedures as the norm group. (The assumption is that they have been taught similar materials.)"[3]

---

[3] Kwalwasser-Ruch Test of Music Accomplishment Manual, Bureau of Educational Research, University of Iowa, 1927.

Table 1[4-5]

Reliability Coefficients

	SUB TEST	$r_{\frac{1}{2}\frac{1}{2}}$	r12
1.	Knowledge of Musical Symbols and Terms	.85	.97
2.	Recognition of Syllable Names	.77	.87
3.	Detection of Pitch Errors in a Familiar Melody	.63	.77
4.	Detection of Time Errors in a Familiar Melody	.56	.72
5.	Recognition of Pitch Names	.81	.89
6.	Knowledge of Time Signatures	.67	.80
7.	Knowledge of Key Signatures	.91	.95
8.	Knowledge of Note Values	.54	.70
9.	Knowledge of Rest Values	.56	.72
10.	Recognition of Familiar Melodies from Notation	.63	.77
	Total Scores (ten tests)	.95	.97
	Standard Deviation		51.5
	Probable Error		5.9
	Mean		156.0

---

[4] The column headed $r_{\frac{1}{2}\frac{1}{2}}$ was obtained by correlation of the odd-numbered items against the even-numbered items. The column headed r 12 was obtained from the preceding column by means of the Spearman-Brown prophecy formula.

[5] Kwalwasser, Jacob & Ruch, G.M., Manual of Directions for Kwalwasser-Ruch Test of Musical Accomplishment for Grades IV-XII, 1927, Bureau of Educational Research and Service, State University of Iowa.

Table 2[6]

Percentile Scores by Grades Based Upon 5,414 pupils.

Percentile	GRADES					
	IV	V	VI	VII	VIII	IX-XII
90%-ile	107	128	153	168	191	219
80%-ile	95	111	134	152	174	200
70%-ile	82	95	117	132	153	186
60%-ile	74	80	108	121	142	172
50%-ile(Median)	66	74	96	106	127	156
40%-ile	58	67	82	94	116	139
30%-ile	50	60	67	83	106	123
20%-ile	42	50	55	68	93	102
10%-ile	30	40	48	54	72	78
Number	634	713	1487	1128	672	780

[6] Kwalwasser, Jacob and Ruch, G.M., Manual of Directions for Kwalwasser-Ruch Test of Musical Accomplishment for Grades IV-XII, Bureau of Research and Service, State University of Iowa, 1927.

BIBLIOGRAPHY

Broom, M.E., Educational Measurements in Elementary School, Mc Graw-Hill Book Company, New York, 1939.

Brooks, B. Marion and Brown, Harry A., Music Education in the Elementary School, American Book Company, New York, 1946.

Dykema, Peter W. and Gehrkens, Karl W., The Teaching and Administration of High School Music, C.C. Birchard Company, Boston, 1941.

Gehrkens, Karl W., Music Tests: Pros and Cons, Musician, March,1938.

Kwalwasser, Jacob, Tests and Measurements in Music, C.C. Birchard and Company, Boston, 1927.

Kwalwasser, Jacob, Problems in Public School Music, M. Witmark and Sons, New York City, N.Y., 1932.

Kwalwasser, Jacob, Tests and Measurements in Music, Psychological Bulletin, 1925, pp. 284-301.

Lincoln, Edward A. and Workman, Linwood L., Testing and Use of Test Results, The MacMillan Company, 1936, N.Y.

Mursell, J.L., Measuring Musical Ability and Achievement. A Study of the Correlations of Seashore Test Scores and Other Variables. Journal of Educational Research, 25: 116-26.

Schoen, Max ,Report of the Committee on Music Tests and Measurements, 1934 and 1936 Proceedings of the National Association of Music Teachers.

Walker, Helen M., Elementary Statistics, Henry Holt and Company, New York, N.Y., 1943.

BOSTON UNIVERSITY
1 1719 02551 0431

ACCOPRESS BINDER

[illegible]

MADE BY
ACCO PRODUCTS, INC.
OGDENSBURG, N. Y., U. S. A.